Application of a Semiparametric Statistical Method to Spacecraft Sensor and Calibration Data

Benjamin Kedem*, Julie Thienel† and Rick Harman‡

A semiparametric statistical method is applied to spacecraft sensor data and spacecraft sensor residual data. The method stipulates a reference distribution and deviations from it. The problem is to estimate the reference distribution and the distorted distributions from all available data from all the sources under consideration. The problem can be solved by assuming a distortion form and independent data.

INTRODUCTION

Suppose that m instruments $I_1, ..., I_q, I_m$ all measure the same or related quantities and that I_m is chosen as the "reference instrument". The other instruments are thought of as "distortions" in the sense that the probability distributions of their data are distortions of the probability distribution of the reference data. The problem is to *combine* the information from all the instruments in order to construct an improved estimate of the reference (true) probability distribution and of the deviant distributions.

The instruments we have in mind include, but certainly are not limited to, sensors such as gyroscopes, star trackers, sun sensors, earth sensors, and other attitude and navigation sensors. The method can be applied to space borne instruments where ground truth is taken as the reference, or to residual data obtained after onboard sensors are calibrated. In this work, the method is applied to measurements from

*Department of Mathematics, University of Maryland, College Park, MD 20742, phone: 301-405-5061/5112, fax: 301-314-0827, email: bnk@math.umd.edu

†Flight Dynamics Analysis Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, phone: 301-286-9033, fax: 301-286-0369, email: julie.thienel@nasa.gov

‡Flight Dynamics Analysis Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, phone: 301-286-5125, fax: 301-286-0369, email: richard.r.harman@nasa.gov

the Rossi X-Ray Timing Explorer (RXTE) star trackers, and residual data from the calibrated Upper Atmosphere Research Satellite (UARS) earth sensors.

The present work is based on Fokianos, Kedem, Qin, Haferman, and Short (Ref. 1), Fokianos, Kedem, Qin, Short (Ref. 2), and Kedem, Wolff, and Fokianos (Ref. 3).

STATISTICAL FORMULATION

We describe briefly the main statistical ideas and introduce our notation. Except for an assumed distortion form and independent data, nothing else is assumed.

Let $\mathbf{x}_j = (x_{j1}, \ldots, x_{jn_j})'$ represent independent observations from instrument I_j, where x_{ji} has an *unknown* probability density $g_j(x)$. Then, holding $g \equiv g_m$ as a reference, again unknown, we assume that each of the $q = m-1$ probability densities $g_1(x), ..., g_q(x)$ is a tilted form of $g(x)$,

$$g_j(x) = \exp(\alpha_j + \beta_j h(x))g(x),\ j = 1, ..., q \tag{1}$$

where α_j depends on β_j. Quite a few distributions satisfy (1) including the normal, lognormal, and gamma distributions.[4] More precisely, in the normal case when each datum has the same variance, the regression function is $h(x) = x$, whereas for the corresponding lognormal and for certain gamma populations $h(x) = \log(x)$. Another example is provided by multinomial logistic regression.[2] Distortions such as (1) have been studied in Refs. 5 and 6, while the validation problem of model (1) has been studied in Ref. 7 for the special case $m = 2$ using a generalized moments specification test.

When $\beta_1 = \cdots = \beta_q = 0$, also $\alpha_1 = \cdots = \alpha_q = 0$, then there are no distortions, that is, all the sensors give the same identically distributed data and we have $g_1(x) = g_2(x) = \cdots = g_q(x) = g(x)$. When some β_j are not equal to 0, the sensors deviate statistically from the benchmark or reference sensor. Thus, $\beta_j \neq 0$ points to a difference between I_j and the reference sensor. In the present context, calibration of I_j means that $g_j(x)$ must be multiplied by $\exp(-\alpha_j - \beta_j h(x))$ to increase its reliability.

Let $n = n_1 + \cdots + n_q + n_m$, where n_j is the size of x_j, and $\rho_j = n_j/n_m$, $j = 1, ..., q$. Combine all the data in the vector $\mathbf{t} = (\mathbf{x}_1', ..., \mathbf{x}_q', \mathbf{x}_m')' = (t_1, ..., t_n)'$. Interestingly, the relative sample sizes ρ_j play an important role in estimation and hypothesis testing. It is also convenient to introduce the notation,

$$w_j(t) = \exp(\alpha_j + \beta_j h(t)),\ j = 1, ..., q.$$

Estimation

Using the combined data $\mathbf{t} = (t_1, ..., t_n)'$, the semiparametric estimation of the reference $g(x)$ and all the distortion parameters $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_q)'$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)'$,

and the hypothesis test that there are no distortions $\mathrm{H}_0 : \beta_1 = \cdots = \beta_q = 0$, have been undertaken in Ref. 2 using *empirical likelihood*[8] which leads to surprisingly simple procedures that are useful and easy to implement.

The reference probability density (pdf) $g(x)$ is estimated by a discrete probability distribution $\hat{p}_i = \hat{p}(t_i)$ at every t_i, $i = 1, ..., n$. The rationale behind this is that we really are estimating the corresponding cumulative distribution function (cdf)

$$G(x) = \int_{-\infty}^{x} g(u)du$$

by a step function $\hat{G}(x)$ with jumps at the t_i's, and $\hat{p}_i$ is the corresponding probability mass function.

Notice that $g(x) = g_m(x)$ is estimated from the combined data $\mathbf{t}$ and not just from the mth sample $x_{m1}, ..., x_{mn_m}$.

A maximum likelihood estimator of $G(x)$ can be obtained by maximizing the likelihood over the class of step cdf's with jumps at the observed values $t_1, ..., t_n$. Accordingly, if $p_i = dG(t_i)$, $i = 1, .., n$, the likelihood becomes

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp(\alpha_1 + \beta_1 h(x_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\alpha_q + \beta_q h(x_{qj})) \tag{2}$$

By maximizing (2) with respect to the p_i subject to the constraints that the p_i and all the corresponding distortions sum up to 1 we obtain the estimates

$$\hat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 \exp(\hat{\alpha}_1 + \hat{\beta}_1 h(t_i)) + \cdots + \rho_q \exp(\hat{\alpha}_q + \hat{\beta}_q h(t_i))} \tag{3}$$

and therefore, with $I(B)$ the indicator of the event B,

$$\hat{G}(t) = \frac{1}{n_m} \cdot \sum_{i=1}^{n} \frac{I(t_i \leq t)}{1 + \rho_1 \exp(\hat{\alpha}_1 + \hat{\beta}_1 h(t_i)) + \cdots + \rho_q \exp(\hat{\alpha}_q + \hat{\beta}_q h(t_i))} \tag{4}$$

The estimates of the α_j and β_j are solutions of score equations in terms of a profile log-likelihood l (see Ref. 2) for $j = 1, ..., q$,

$$\begin{aligned} \frac{\partial l}{\partial \alpha_j} &= -\sum_{i=1}^{n} \frac{\rho_j w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + n_j = 0 \\ \frac{\partial l}{\partial \beta_j} &= -\sum_{i=1}^{n} \frac{\rho_j h(t_i) w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + \sum_{i=1}^{n_j} h(x_{ji}) = 0 \end{aligned} \tag{5}$$

The solutions $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, ..., \hat{\alpha}_q)'$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_q)'$, are known to have an asymptotic normal distribution with mean $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and a $2q \times 2q$ covariance matrix $\boldsymbol{\Sigma}/n$, where

$$\boldsymbol{\Sigma} = \mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1} \tag{6}$$

and

$$\mathbf{V} \equiv Var\left[\frac{1}{\sqrt{n}}\nabla l(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]$$

$$-\frac{1}{n}\nabla\nabla' l(\boldsymbol{\alpha}, \boldsymbol{\beta}) \to \mathbf{S}, \quad n \to \infty.$$

Hypothesis Testing

The hypothesis that all the probability distributions corresponding to the different instruments are the same,

$$\mathrm{H}_0 : \beta_1 = \cdots = \beta_\mathrm{q} = 0$$

can be tested by a statistic which depends on n, the variance of $h(t)$ with respect to the reference distribution $g(x)$ which we can approximate with the help of (3), $\hat{\boldsymbol{\beta}}$, and the sample size ratios ρ_j.

Let $\mathbf{A}_{11}$ be a $q \times q$ matrix whose jth diagonal element is

$$\frac{\rho_j[1 + \sum_{k\neq j}^{q} \rho_k]}{[1 + \sum_{k=1}^{q} \rho_k]^2}$$

and otherwise for $j \neq j'$, the jj' element is

$$\frac{-\rho_j \rho_{j'}}{[1 + \sum_{k=1}^{q} \rho_k]^2}$$

Then, under H_0,

$$\mathcal{X}_1 = n\mathrm{Var}[h(t)]\hat{\boldsymbol{\beta}}'\mathbf{A}_{11}\hat{\boldsymbol{\beta}} \tag{7}$$

is approximately distributed as $\chi^2(q)$, and H_0 is rejected for large values of $\mathcal{X}_1 = n\mathrm{Var}[h(t)]\hat{\boldsymbol{\beta}}'\mathbf{A}_{11}\hat{\boldsymbol{\beta}}$.

In practice, the $\mathrm{Var}[h(t)]$ needed for evaluating $\mathcal{X}_1$ is estimated from

$$\sum_{i=1}^{n} h^2(t_i)\hat{p}_i - \left(\sum_{i=1}^{n} h(t_i)\hat{p}_i\right)^2. \tag{8}$$

Simulations indicate that under normality, the test based on $\mathcal{X}_1$ is not dominated by the common F-test, while it is more powerful than the F-test for non-normal data, provided $h(x)$ is known or chosen wisely. Recall that for normal data with equal variance $h(x) = x$, and for certain skewed data, $h(x) = \log(x)$ is the proper choice. The general linear hypothesis is discussed in Ref. 2.

ILLUSTRATION OF THE METHOD

The method is useful when we wish to detect differences or study similarities between probability distributions. Experience shows that the method is quite sensitive and it can detect even small distributional differences. It is also useful for estimating probabilities from the combined data obtained from many sources. This is illustrated in terms of simulated and real data.

Simulated Normal Data

To get an idea as to the sensitivity of the method we employ simulated normal (Gaussian) data with $h(x) = x$, $m = 3$, and sample sizes $n_1 = 200, n_2 = 300, n_3 = 250$, for $N(\mu_1, 1)$, $N(\mu_2, 1)$, $N(0, 1)$, holding $N(0, 1)$ as the reference and varying μ_1, μ_2. Greater values of μ_1, μ_2 result in a greater distortion. From Table 1, very small deviations or distortions are more difficult to detect as expressed by a relatively large p-value, say greater than 0.05. The situation improves greatly as the distortion becomes more pronounced leading to small p-values, much smaller than the nominal 0.05, or equivalently larger values of $\mathcal{X}_1$.

Table 1 Sensitivity of the semiparametric method measured by p-value for $m = 3, q = 2$ and normal data with different means μ_1, μ_2 and variance 1. The reference distribution is $N(0, 1)$.

μ_1, μ_2	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\mathcal{X}_1$	p-value
0.01, 0.00	0.000	-0.008	-0.002	-0.063	0.640	0.726098
0.01, 0.05	0.001	0.064	-0.002	0.133	2.554	0.278895
-0.10, 0.05	-0.008	-0.105	0.000	0.005	3.052	0.212731
-0.10, 0.10	-0.002	-0.040	-0.008	0.180	7.645	0.021869
-0.10, 0.20	-0.009	-0.138	-0.027	0.230	17.406	0.000166
0.20, 0.30	-0.002	0.073	-0.073	0.387	24.332	0.000005
1.00, 0.10	-0.470	0.996	-0.008	0.182	127.764	0.000000

It is important to note that the choice $h(x) = x$ is appropriate when the variances of the various data sets are not drastically different, but when the latter is the case, $h(x) = x^2$ may be appropriate. This is exactly the case for Gaussian (normal) data sets with substantially different variances. To illustrate this, consider the case of Gaussian data $N(0, 2)$, $N(0, 4)$, $N(0, 1)$ where $N(0, 1)$ is the reference pdf. Let the sample sizes be $200, 300, 250$, respectively. Then with $h(x) = x^2$, we obtained

$$(\alpha_1, \beta_1, \alpha_2, \beta_2) = (-0.367, 0.258, -0.628, 0.353)$$

with a very small p-value of 1.739686e-08, as it should since the variances are very different (i.e. "different sensors"). On the other hand, with $h(x) = x$,

$$(\alpha_1, \beta_1, \alpha_2, \beta_2) = (-0.006, -0.048, -0.002, 0.077)$$

with a much larger p-value of 0.09459638 from which we may conclude erroneously at level 0.05 that the sensors are "the same" when they are not.

It is interesting to compare the theoretical SE's obtained from (6) and those given explicitly in Ref. 2 with simulation results. Again consider the case $N(0, 2)$, $N(0, 4)$ and $N(0, 1)$ with respective sample sizes $200, 300, 250$, and holding $N(0, 1)$ as reference. The SE's results with 200 runs which gave 200 estimates for each of $\hat{\alpha}_1$,$\hat{\beta}_1$,$\hat{\alpha}_2$,$\hat{\beta}_2$ are given in Table 2. There the estimated standard error of $\hat{\alpha}_1$ is the sample SE obtained from 200 values of $\hat{\alpha}_1$, etc. We see that the theoretical ("Theory") and estimated ("Est.") SE's are fairly close. Similar close results are obtained with $h(x) = \log(x)$ for lognormal data.

Table 2 Comparison of theoretical versus estimated SE's of $\hat{\alpha}_1$,$\hat{\beta}_1$,$\hat{\alpha}_2$,$\hat{\beta}_2$ for $h(x) = x^2$ using $N(0, 2)$, $N(0, 4)$, $N(0, 1)$ data.

SE($\hat{\alpha}_1$)		SE($\hat{\beta}_1$)		SE($\hat{\alpha}_2$)		SE($\hat{\beta}_2$)	
Theory	Est.	Theory	Est.	Theory	Est.	Theory	Est.
0.069	0.069	0.056	0.054	0.078	0.075	0.056	0.052

Sensor Data

We apply the semiparametric method to the x-coordinates of two star trackers, denoted as star1 and star2, to study how different the corresponding probability distributions are. Two random samples of size 500 each were obtained from each instrument and the corresponding probability distributions (3) were estimated from the *combined* data. Holding star2 as the reference, the results with $h(x) = x$ are given in Table 3. Certainly, the two data sets are very different as expressed by very large parameter values and a zero p-value. However, when the samples come from the same instrument the parameter estimates and their standard errors decrease and the p-values increase dramatically. Thus, judging by the p-values, the method identifies the samples correctly.

Next we compare in Figures 1 and 2 the estimated probability densities with each other and with the corresponding histograms. It should be emphasized that both pdf's were obtained from the combined data set whereas each histogram is from a particular single set, either from star1 or star2, but not both. Figure 2 illustrates clearly the difference between the pdf's.

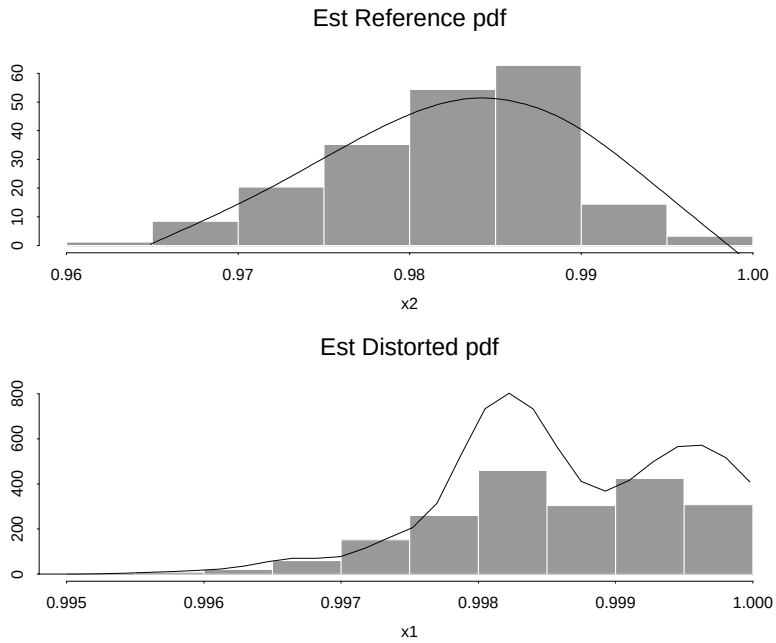

Figure 1 Estimated reference pdf (top) and its distortion (bottom) and the corresponding histograms. $h(x) = x$

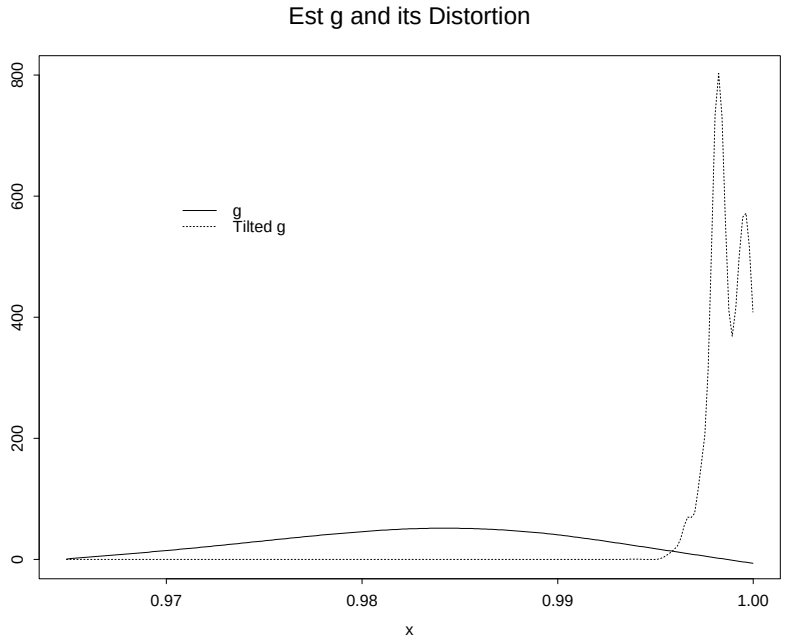

Figure 2 Estimated reference pdf, $g(x)$, and its distortion for the star tracker data.

Table 3 Sensitivity of the semiparametric method measured by p-value for $m = 2, q = 1$ for star trackers star1 and star2, holding star2 as reference. The values in parentheses are standard errors obtained from (6).

Source1	Source2	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
Star1	Star2	-2438.56	2450.71	55906.500	0.000
		(599.43)	(602.11)		
Star1	Star1	20.78	-20.81	0.090	0.762
		(68.26)	(68.36)		
Star2	Star2	-0.92	0.94	0.009	0.924
		(9.69)	(9.86)		

The estimated reference cdf $\hat{G}(x)$ belonging to star2 and given in (4) is shown in Figure 3. From $\hat{G}(x)$, the probability that an x-coordinate value is less that 0.98 is $\hat{G}(0.98) = 0.326$, and the probability of a value falling between 0.98 and 0.99 is $\hat{G}(0.99) - \hat{G}(0.98) = 0.586$, and so on. These probabilities for star1 are practically zero: $\hat{G}_1(0.98) = 1.15 \times 10^{-30}$, and $\hat{G}_1(0.99) - \hat{G}_1(0.98) = 1.28 \times 10^{-11}$, and so on. This can also be seen from Figure 2. Notice that $\hat{G}_1(x)$ is obtained by multiplying (3) by the distortion and proceeding as in (4) with the distorted cdf replacing the reference cdf.

Our next example deals with the residual data from the x-axis for the two UARS earth sensors referred to as "esa1" and "esa2". The residual is computed as the difference between the measured Earth vector and the expected Earth vector, both in the spacecraft body coordinates. More precisely, the residual is the result of calibration, the process which estimates all the errors in the measured vector and then removes them. The residual is what is left after all the calibration parameters are removed. If the calibration of the sensors is perfect, and the vectors are not corrupted by any noise, each data set would be zero. If the calibration is perfect, the data would be "white" noise, with zero mean.

Teting for white noise refers to a joint distribution property, while our method produces inference about marginal distributions when the data comes from many sources. Thus our method is not suitable for white noise testing per se. Still, it is interesting to measure the difference between the distributions of the two "white" noise sources and to get an idea of how *probable* it is to realize values around zero in the two cases. This, however, is accomplished with our method. Thus taking esa2 as the reference instrument, and using $h(x) = x$, we have $(\hat{\alpha}_1, \hat{\beta}_1) = (-0.149, 489.395)$ with corresponding standard errors $(0.026, 66.612)$ and a p-value of 6×10^{-15}. The estimated reference pdf (the estimated pdf of data from esa2) and its distortion (the estimated pdf of data from esa1) are shown in Figure 4. Our analysis shows that these two distributions are very different and yet both are supported on intervals about the zero value. *That is, the probability of a residual being between -0.002 and*

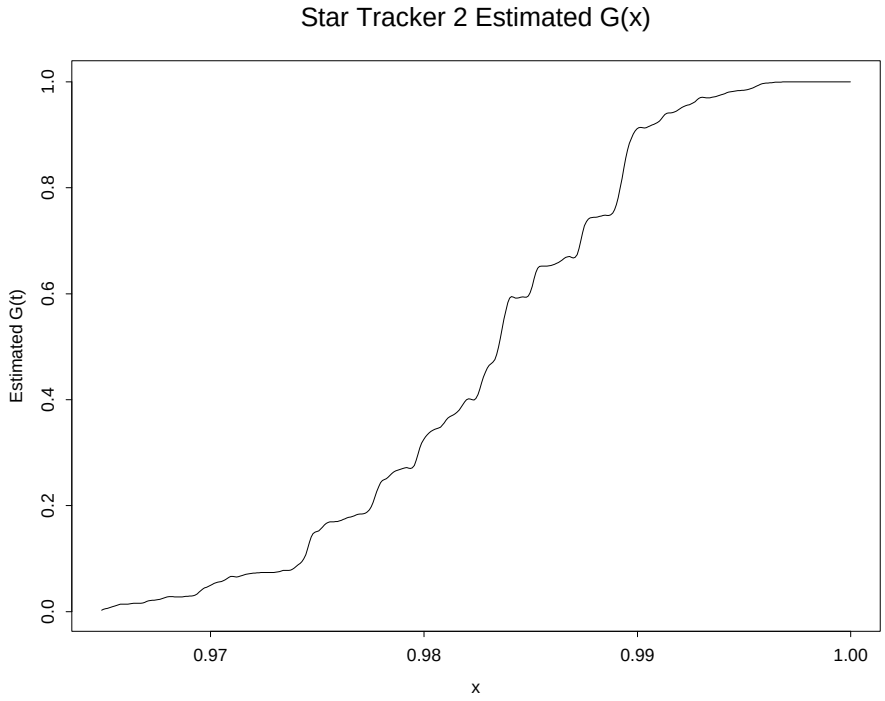

Figure 3 Estimated reference cdf, $G(x)$, of star tracker 2 when $h(x) = x$.

0.003 is nearly 1, and this probability was estimated from the combined residual data from both esa1 and esa2.

SUMMARY

A semiparametric statistical method based on the empirical likelihood (2) was applied to spacecraft sensor data. The method combines the data from all the sources in estimating the probability distribution associated with each data source. The remarkable fact is that probability distributions are derived as a result of an optimization problem assuming a distortion or tilt form and independent data. Nothing else is assumed. The form of the distortion may be generalized to the case where β_1 is a vector and $h(x)$ is vector valued. This generalization is studied elsewhere.

The preliminary results here were limited to two different sensors, and different types of data, namely observed star data and Earth sensor residual data. In the future, this method could be applied to sensor data to determine deviations between identical sensors, to give an indication of an impending sensor failure, for example. Also, additional residual data could be tested to determine the effectiveness of a calibration procedure. Additionally, the method could be applied to navigation data

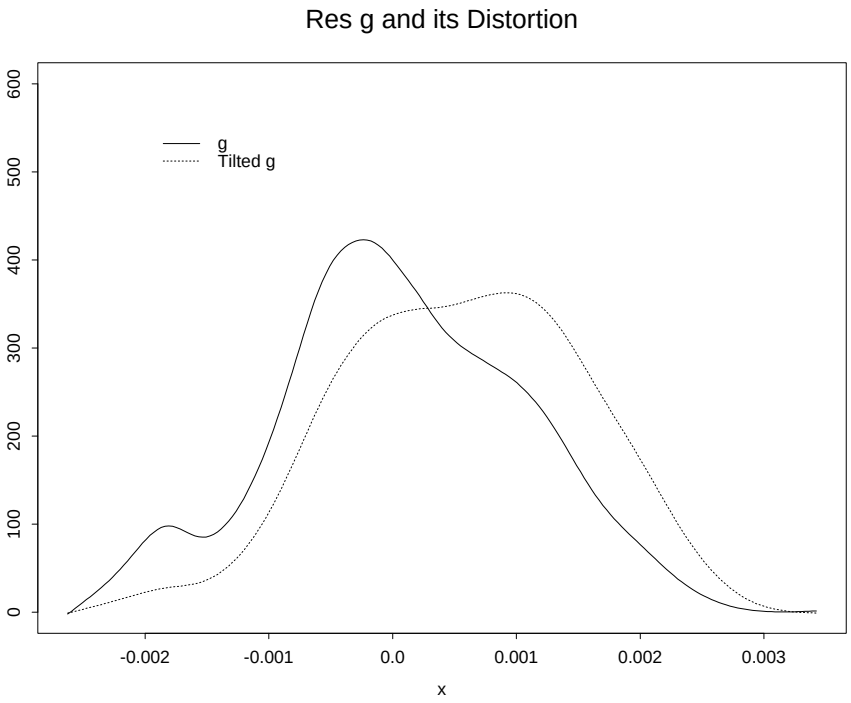

Figure 4 Estimated reference pdf $g(x)$ and its distortion for the UARS data. $h(x) = x$.

to determine the statistical properties of transmitted ranging data, or data processed by onboard receivers.

REFERENCES

1. Fokianos, K., Kedem, B., Qin, J., Haferman, J., and Short, D., "On Combining Instruments," *Journal of Applied Meteorology*, Vol. 37, 1998, pp. 220–226.

2. Fokianos, K., Kedem, B., Qin, J., and Short, D. A., "A semiparametric approach to the one-way layout," *Technometrics*, Vol. 43, 2001, pp. 56–65.

3. Kedem, B., Wolff, D., and Fokianos, K., "Statistical Comparison of Algorithms," submitted.

4. Kay, R. and (1987), S. L., "Transformations of the explanatory variables in the logistic regression model for binary data," *Biometrika*, Vol. 74, 1987, pp. 495–501.

5. Efron, B. and Tibshirani, R., "Using specially designed exponential families for density estimation," *The Annals of Statistics*, Vol. 24, 1996, pp. 2431–2461.

6. Qin, J. and Zhang, B., "A goodness of fit test for logistic regression models based on case-control data," *Biometrika*, Vol. 84, 1997, pp. 609–618.

7. Fokianos, K., Peng, A., and Qin, J., "A generalized–moments specification test for the logistic link," *The Canadian Journal of Statistics*, Vol. 27, 1999, pp. 735–750.

8. Owen, A. B., *Empirical Likelihood*, Chapman and Hall/CRC, New York, 2001.